



Revista Hambatu Science

Ambato – Ecuador / ISSN 3151-815X (en línea) / julio - septiembre 2026

Volumen 1, Número 3

<https://doi.org/10.63862/rhs-v1n3-143-161-2026>

El vínculo simbiótico entre inteligencia artificial y evaluación educativa: principios éticos, sesgos, derechos humanos y bienestar emocional del estudiante

The symbiotic link between artificial intelligence and educational assessment: ethical principles, bias, human rights and students' emotional well-being

Celina Hernández Armenta
Universidad Nacional Autónoma de México

Edmundo Cervantes Meza
Benemérita Universidad Autónoma de Puebla

Guadalupe Patricia Alegre Campos
Universidad Nacional Rosario Castellanos

Yolanda del Sol Ortega Cruz
Universidad Autónoma de Zacatecas

Angelica Rosendo Olivares
Universidad Nacional Autónoma de México



DOI: <https://doi.org/10.63862/rhs-v1n2-143-161-2026>

El vínculo simbiótico entre inteligencia artificial y evaluación educativa: principios éticos, sesgos, derechos humanos y bienestar emocional del estudiante

Celina Hernández Armenta

Universidad Nacional Autónoma de México

nalice1987@hotmail.com

<https://orcid.org/0009-0000-3098-4184>

Ciudad de México, México

Yolanda del Sol Ortega Cruz

Universidad Autónoma de Zacatecas

sol.ortega@uaz.edu.mx

<https://orcid.org/0009-0007-5800-0715>

Zacatecas, México

Edmundo Cervantes Meza

Benemérita Universidad Autónoma de Puebla

e-di-13@hotmail.com

<https://orcid.org/0009-0009-5536-0904>

Puebla, México

Angelica Rosendo Olivares

Universidad Nacional Autónoma de México

angelitaups@gmail.com

<https://orcid.org/0009-0007-5779-762X>

Ciudad de México, México

Guadalupe Patricia Alegre Campos

Universidad Nacional Rosario Castellanos

az09.alejana@gmail.com

<https://orcid.org/0009-0001-2407-2864>

Ciudad de México, México

Recibido: 2026-06-26

Aceptado: 2026-06-29

Publicado: 2026-07-08

Resumen

El presente artículo analiza la pertinencia ética, jurídica y psicoeducativa del uso de la inteligencia artificial en la evaluación del aprendizaje. A partir de una investigación teórico-documental, se revisan fuentes académicas, normativas e institucionales sobre inteligencia artificial educativa, evaluación formativa, sesgos algorítmicos, derechos humanos, protección de datos y bienestar emocional del estudiante. El objetivo es proponer criterios mínimos para que la evaluación asistida por inteligencia artificial funcione como herramienta de apoyo docente y no como mecanismo opaco de clasificación, vigilancia o exclusión. Los resultados del análisis muestran que el vínculo entre inteligencia artificial y evaluación educativa es simbiótico: la evaluación produce datos que alimentan sistemas automatizados, mientras que estos sistemas reconfiguran la forma de observar, interpretar y decidir sobre el aprendizaje. Se identifican riesgos asociados a sesgos, opacidad, afectaciones a la privacidad, presión emocional, dependencia tecnológica y debilitamiento del juicio pedagógico. Como aportación, se propone una matriz de principios éticos organizada en proporcionalidad, transparencia, explicabilidad, revisión humana significativa, equidad, protección de datos, inclusión y cuidado emocional. Se concluye que la inteligencia artificial puede fortalecer la evaluación solo si permanece subordinada a la dignidad humana, la justicia educativa y la responsabilidad institucional.

Palabras clave: inteligencia artificial educativa, evaluación del aprendizaje, derechos humanos, sesgos algorítmicos, bienestar emocional, ética educativa, supervisión humana.

The symbiotic link between artificial intelligence and educational assessment: ethical principles, bias, human rights and students' emotional well-being

Abstract

This article analyzes the ethical, legal, and psychoeducational relevance of using artificial intelligence in learning assessment. Based on a theoretical-documentary research design, the study reviews academic, normative, and institutional sources on educational artificial intelligence, formative assessment, algorithmic bias, human rights, data protection, and students' emotional well-being. The objective is to propose minimum criteria so that AI-assisted assessment operates as a tool that supports teaching rather than as an opaque mechanism of classification, surveillance, or exclusion. The analysis shows that the relationship between artificial intelligence and educational assessment is symbiotic: assessment generates data that feed automated systems, while these systems reshape the ways in which learning is observed, interpreted, and decided upon. The study identifies risks related to bias, opacity, privacy harms, emotional pressure, technological dependence, and the weakening of pedagogical judgment. As a contribution, the article proposes a matrix of ethical principles organized around proportionality, transparency, explainability, meaningful human review, equity, data protection, inclusion, and emotional care. It concludes that artificial intelligence can strengthen educational assessment only if it remains subordinate to human dignity, educational justice, and institutional responsibility.

Keywords: educational artificial intelligence, learning assessment, human rights, algorithmic bias, emotional well-being, educational ethics, human oversight.

Introducción

La incorporación de la inteligencia artificial en los procesos educativos ha dejado de ser una discusión prospectiva para convertirse en una práctica cotidiana en instituciones de educación básica, media superior y superior. Plataformas adaptativas, sistemas de retroalimentación automática, asistentes generativos, detectores de similitud textual, analíticas de aprendizaje y herramientas de proctoring ya participan en actividades que antes dependían casi exclusivamente de la interacción docente-estudiante. En este contexto, la evaluación del aprendizaje se ubica en una zona especialmente sensible, porque no solo mide resultados, sino que clasifica desempeños, orienta trayectorias, condiciona oportunidades académicas y puede incidir en becas, acreditaciones, permanencia escolar o reputación estudiantil (Holmes et al., 2019; Miao & Holmes, 2023).

El atractivo de la inteligencia artificial aplicada a la evaluación educativa descansa en su promesa de rapidez, personalización, seguimiento continuo y toma de decisiones basada en datos. La tecnología puede apoyar la elaboración de instrumentos, detectar rezagos tempranos, generar retroalimentación inmediata y ofrecer información útil para ajustar estrategias didácticas. Sin embargo, esas ventajas no eliminan la necesidad de un juicio pedagógico situado. Evaluar no consiste únicamente en procesar respuestas, asignar puntajes o generar predicciones; implica interpretar procesos de aprendizaje, reconocer contextos, identificar barreras, acompañar trayectorias y sostener un vínculo formativo que difícilmente puede reducirse a una operación algorítmica (Black & Wiliam, 1998; Nicol & Macfarlane-Dick, 2006).

El problema de investigación surge cuando la eficiencia tecnológica se convierte en criterio suficiente de legitimidad. Un sistema automatizado puede ser rápido y, al mismo tiempo, injusto; puede ser preciso en términos estadísticos y, aun así, producir efectos discriminatorios; puede generar información abundante, pero incomprensible para docentes y estudiantes. En educación, la neutralidad aparente de la tecnología puede ocultar decisiones normativas sobre qué se valora, qué se mide, qué se considera error, qué estilos de aprendizaje se reconocen y qué formas de expresión quedan penalizadas (Baker & Hawn, 2022; Selwyn, 2019).

Este artículo sostiene que entre inteligencia artificial y evaluación educativa se configura un vínculo simbiótico. La evaluación produce datos, criterios, indicadores y patrones que alimentan a los sistemas automatizados; a su vez, la inteligencia artificial reorganiza la manera en que las instituciones observan, interpretan y deciden sobre el aprendizaje. Esta relación puede fortalecer la justicia educativa cuando se gobierna bajo principios de derechos humanos, equidad y supervisión humana; pero también puede profundizar desigualdades cuando opera como caja negra, cuando clasifica sin contexto o cuando sustituye indebidamente la responsabilidad docente e institucional (OECD, 2023; UNESCO, 2022).

La pregunta orientadora es la siguiente: ¿cuáles son los principales riesgos éticos, jurídicos y psicoemocionales del uso de inteligencia artificial en la evaluación del aprendizaje, y qué criterios deberían orientar su implementación responsable? El objetivo general consiste en analizar dichos riesgos y proponer una matriz de principios para una evaluación asistida por inteligencia artificial compatible con derechos humanos, justicia educativa y bienestar emocional del estudiante. Como objetivos específicos, se busca: a) examinar el papel de la inteligencia artificial en la evaluación del aprendizaje; b) identificar riesgos de sesgo, opacidad, privacidad y daño emocional; c) articular principios éticos y jurídicos aplicables; y d) proponer criterios institucionales de implementación responsable.

La contribución del trabajo radica en vincular tres campos que suelen estudiarse por separado: la ética de la inteligencia artificial, la teoría de la evaluación educativa y la protección de derechos humanos en contextos escolares. Con ello, el artículo busca aportar una lectura interdisciplinaria útil para docentes, autoridades educativas, comités de ética, diseñadores de políticas institucionales y personas investigadoras interesadas en la gobernanza responsable de tecnologías emergentes.

Estado del arte

La literatura sobre inteligencia artificial en educación ha crecido de manera acelerada durante la última década. Las revisiones sistemáticas muestran que las aplicaciones más frecuentes se concentran en sistemas tutoriales inteligentes, analíticas de aprendizaje, automatización de retroalimentación, predicción de desempeño y personalización de contenidos (Zawacki-Richter et al., 2019). Estos desarrollos han permitido imaginar modelos de enseñanza más adaptativos,

pero también han abierto interrogantes sobre transparencia, validez, responsabilidad, sesgos y efectos subjetivos en estudiantes y docentes. En ese debate, Popenici y Kerr (2017) advierten que la incorporación de inteligencia artificial en educación superior exige una lectura crítica de sus promesas, porque el aprendizaje no puede reducirse a transferencia de información ni a clasificación automatizada de resultados.

En el campo de la evaluación, la discusión no puede separarse de los debates clásicos sobre validez y función formativa. Messick (1995) sostuvo que la validez de una evaluación no se limita a la calidad técnica de un instrumento, sino que incluye la interpretación de los resultados y sus consecuencias. Este planteamiento resulta decisivo para evaluar sistemas automatizados: no basta con preguntar si un algoritmo calcula de forma consistente, sino si sus inferencias son justas, comprensibles y pedagógicamente pertinentes. Black y Wiliam (1998), así como Nicol y Macfarlane-Dick (2006), han mostrado que la retroalimentación mejora el aprendizaje cuando ayuda al estudiante a comprender su desempeño y a regular sus propias estrategias; por ello, una retroalimentación automatizada que sea opaca, genérica o sancionadora puede perder sentido formativo.

Los documentos internacionales recientes han insistido en que la inteligencia artificial en educación debe orientarse por un enfoque centrado en el ser humano. La UNESCO plantea que el uso de inteligencia artificial generativa en educación e investigación exige preservar la agencia humana, la inclusión, la diversidad cultural y la protección de datos (Miao & Holmes, 2023). Asimismo, la Recomendación sobre la Ética de la Inteligencia Artificial establece principios de proporcionalidad, no daño, seguridad, equidad, privacidad, transparencia, explicabilidad, trazabilidad y supervisión humana (UNESCO, 2022). A este marco se suman las directrices europeas de inteligencia artificial confiable, que enfatizan legalidad, robustez técnica y responsabilidad humana (High-Level Expert Group on Artificial Intelligence, 2019). Estos principios son relevantes para la evaluación porque cualquier decisión automatizada que afecte trayectorias educativas debe poder explicarse, revisarse y corregirse.

La Organisation for Economic Co-operation and Development (OECD, 2023) ha advertido que la digitalización educativa no puede reducirse a infraestructura o adquisición de software, sino que requiere ecosistemas de gobernanza, formación docente, seguridad de datos y evaluación de impactos. De manera complementaria, los marcos de competencias en inteligencia artificial

para docentes y estudiantes propuestos por la UNESCO subrayan que la alfabetización en IA no debe limitarse al uso instrumental de herramientas, sino incluir comprensión crítica, ética, responsabilidad social y capacidad para reconocer limitaciones y sesgos (UNESCO, 2024a, 2024b).

El riesgo de sesgo algorítmico constituye uno de los temas más desarrollados en la literatura. Baker y Hawn (2022) explican que los sesgos en educación pueden aparecer desde la definición del problema hasta la selección de datos, el entrenamiento del modelo, el despliegue de la herramienta y la interpretación de resultados. En consecuencia, la injusticia algorítmica no es solo una falla técnica, sino un problema institucional y político. Cuando una plataforma reproduce patrones de desigualdad por género, lengua, discapacidad, origen social o acceso tecnológico, la evaluación puede convertirse en mecanismo de exclusión bajo apariencia de objetividad.

Otra línea relevante se refiere a la ética de las analíticas de aprendizaje. Slade y Prinsloo (2013) sostienen que el uso de datos educativos debe considerar consentimiento, transparencia, finalidad, proporcionalidad y responsabilidad. Desde esta perspectiva, el estudiante no debe ser tratado como simple fuente de datos institucionales. Las plataformas evaluativas procesan información académica, conductual y, en ocasiones, emocional; por ello, su uso exige límites claros sobre recolección, conservación, tratamiento y comunicación de resultados.

Finalmente, la dimensión emocional ha cobrado importancia frente al aumento de tecnologías de vigilancia, seguimiento y evaluación permanente. La evaluación siempre ha tenido efectos subjetivos, pero la automatización puede intensificar la sensación de control, la ansiedad por el rendimiento y la desconfianza cuando el estudiante percibe que una máquina decide sin comprender su contexto. La teoría de la autodeterminación muestra que la motivación depende de condiciones de autonomía, competencia y relación; por tanto, una evaluación tecnológica que reduzca la autonomía o debilite la confianza puede afectar la disposición para aprender (Deci & Ryan, 2000; Tang et al., 2023).

El vacío que justifica este artículo consiste en que buena parte de la discusión sobre inteligencia artificial educativa se concentra en beneficios técnicos o en riesgos generales, pero no siempre articula evaluación, derechos humanos y bienestar emocional en una misma matriz operativa.

Por ello, el presente trabajo propone ordenar los riesgos y principios aplicables a partir del vínculo simbiótico entre inteligencia artificial y evaluación educativa.

Metodología

La investigación adopta un enfoque cualitativo de carácter teórico-documental, con alcance analítico y propositivo. Este diseño resulta pertinente porque el objeto de estudio no consiste en medir percepciones de una muestra específica, sino en examinar críticamente fuentes académicas, normativas e institucionales para construir una propuesta de principios aplicables a la evaluación educativa asistida por inteligencia artificial. El análisis documental permite identificar convergencias, tensiones y vacíos entre los discursos pedagógicos, jurídicos y éticos que regulan o problematizan el uso de estas tecnologías.

El corpus se integró mediante criterios de pertinencia temática, actualidad y autoridad de las fuentes. Se consideraron documentos internacionales sobre ética de la inteligencia artificial, educación y derechos humanos; literatura académica sobre inteligencia artificial educativa, evaluación formativa, sesgos algorítmicos y analíticas de aprendizaje; así como marcos institucionales recientes sobre gobernanza, competencias digitales y sistemas de alto riesgo. De manera particular, se revisaron fuentes de UNESCO, OECD, UNICEF, el National Institute of Standards and Technology (NIST, 2023), la Unión Europea y Naciones Unidas, además de investigaciones académicas publicadas en revistas especializadas y editoriales académicas.

El procedimiento de análisis se organizó en cuatro etapas. Primero, se identificaron las funciones principales de la inteligencia artificial en la evaluación educativa: automatización, predicción, personalización, retroalimentación, vigilancia y generación de reportes. Segundo, se clasificaron los riesgos en dimensiones éticas, jurídicas, pedagógicas y psicoemocionales. Tercero, se relacionaron esos riesgos con principios de control institucional: proporcionalidad, transparencia, explicabilidad, trazabilidad, equidad, privacidad, revisión humana y bienestar. Cuarto, se construyó una matriz de implementación responsable orientada a instituciones educativas.

Las categorías analíticas fueron las siguientes: a) función pedagógica de la inteligencia artificial; b) sesgos algorítmicos y desigualdad educativa; c) derechos humanos

comprometidos; d) bienestar emocional del estudiante; e) responsabilidad docente e institucional; y f) principios éticos de implementación. Estas categorías permiten interpretar la evaluación automatizada como un fenómeno sociotécnico, es decir, como una práctica donde interactúan herramientas digitales, decisiones humanas, normas institucionales, datos, contextos de aprendizaje y relaciones de poder.

El estudio tiene una limitación metodológica: al tratarse de una investigación documental, no presenta resultados estadísticos ni generalizables a una población específica. Sin embargo, su fortaleza radica en articular literatura especializada y estándares normativos para formular una propuesta transferible a futuras investigaciones empíricas, lineamientos institucionales o protocolos de evaluación educativa con inteligencia artificial.

Resultados

Los resultados del análisis documental se presentan como hallazgos conceptuales y metodológicos. En primer lugar, se confirma que la inteligencia artificial puede apoyar la evaluación del aprendizaje, pero su legitimidad depende de la finalidad educativa que persigue, del tipo de datos que procesa, de la calidad de sus inferencias, de la posibilidad de revisión humana y del impacto que produce en estudiantes. Por ello, la pregunta central no es si la inteligencia artificial puede evaluar, sino en qué condiciones debe participar en decisiones evaluativas.

En segundo lugar, se identifica que la evaluación asistida por inteligencia artificial genera un ciclo de interdependencia. La institución define qué datos importan, qué resultados espera y qué criterios considera válidos; posteriormente, el sistema automatizado procesa esa información y produce clasificaciones, alertas, sugerencias o retroalimentaciones que modifican la práctica docente. Si ese ciclo no cuenta con transparencia, auditoría y revisión humana, la tecnología puede estabilizar errores o desigualdades bajo apariencia de neutralidad.

En tercer lugar, se observa que los principales riesgos no son independientes entre sí. La opacidad se relaciona con la imposibilidad de impugnar resultados; el sesgo se vincula con desigualdad educativa; la vigilancia constante puede afectar el bienestar emocional; y la falta

de capacitación docente puede convertir las recomendaciones automatizadas en decisiones irreflexivas. En consecuencia, la respuesta institucional debe ser integral y no fragmentaria.

Tabla 1

Riesgos de la evaluación asistida por inteligencia artificial y garantías institucionales mínimas

Riesgo identificado	Efecto posible en el estudiante	Garantía institucional mínima
Sesgo algorítmico	Calificación o clasificación injusta por género, lengua, discapacidad, condición socioeconómica o acceso desigual a tecnología.	Auditoría de sesgos, revisión humana, pruebas con grupos diversos y ajustes razonables.
Opacidad del sistema	Imposibilidad de comprender por qué se obtuvo una calificación, alerta o recomendación.	Explicabilidad, trazabilidad de criterios y comunicación clara de resultados.
Vigilancia excesiva	Ansiedad, autocensura, miedo al error y pérdida de confianza en la institución.	Proporcionalidad, límites de monitoreo, consentimiento informado y evaluación de impacto emocional.
Uso intensivo de datos personales	Afectación a la privacidad y riesgo de usos secundarios no autorizados.	Minimización de datos, finalidad legítima, seguridad, conservación limitada y protección reforzada.
Sustitución del juicio docente	Decisiones rígidas, descontextualizadas o incompatibles con necesidades educativas específicas.	Supervisión humana significativa y posibilidad de corregir o dejar sin efecto recomendaciones automatizadas.
Dependencia tecnológica	Reducción de autonomía, pensamiento crítico y confianza en procesos formativos humanos.	Alfabetización digital crítica, acompañamiento docente y uso auxiliar de la herramienta.

Fuente: elaboración propia con base en Baker y Hawn (2022), Miao y Holmes (2023), OECD (2023), UNESCO (2022, 2024a, 2024b) y UNICEF (2021).

Un cuarto hallazgo es que la evaluación automatizada requiere distinguir entre apoyo pedagógico y decisión institucional. Una herramienta puede sugerir retroalimentación, detectar patrones o advertir posibles rezagos, pero no debe convertirse en autoridad final cuando sus resultados impactan calificaciones, acreditaciones, becas o permanencia escolar. La revisión humana significativa implica que una persona responsable pueda comprender la recomendación, contrastarla con el contexto, escuchar al estudiante y modificar el resultado cuando sea necesario.

Un quinto hallazgo es que los principios éticos no deben permanecer como declaraciones generales, sino convertirse en criterios operativos. La proporcionalidad exige justificar por qué se usa inteligencia artificial y no una medida menos intrusiva. La transparencia obliga a informar qué datos se procesan y con qué finalidad. La explicabilidad requiere que docentes y estudiantes comprendan la lógica general de los resultados. La trazabilidad demanda documentar decisiones, criterios, versiones del sistema y responsables institucionales. La equidad exige evaluar impactos diferenciados. La privacidad impone límites de recolección y conservación de datos. El bienestar emocional obliga a considerar cómo se vive subjetivamente la evaluación.

Tabla 2

Matriz de principios para una evaluación educativa asistida por inteligencia artificial

Principio	Pregunta de control	Aplicación en la evaluación
Finalidad pedagógica legítima	¿El uso de IA responde a una necesidad educativa clara?	Evitar usos meramente punitivos, comerciales o de vigilancia injustificada.
Proporcionalidad y no daño	¿La herramienta es necesaria y no existe una alternativa menos intrusiva?	Usar IA solo cuando aporte valor formativo y no genere daño emocional o discriminación.
Transparencia	¿El estudiante sabe que se usará IA y para qué?	Informar funciones, datos procesados, consecuencias y responsables.
Explicabilidad	¿El resultado puede comprenderse de manera razonable?	Ofrecer razones claras sobre calificaciones, alertas o recomendaciones.
Trazabilidad	¿La institución puede reconstruir cómo se produjo una decisión?	Registrar datos, criterios, responsables, versiones del sistema y revisiones.
Equidad e inclusión	¿La herramienta funciona adecuadamente para grupos diversos?	Auditar sesgos y realizar ajustes razonables para discapacidad, lengua, contexto y brecha digital.
Revisión humana significativa	¿Una persona puede modificar la recomendación automatizada?	Garantizar revisión docente o colegiada antes de decisiones relevantes.
Protección de datos	¿Se recolectan solo datos necesarios y seguros?	Aplicar minimización, seguridad, finalidad, conservación limitada y confidencialidad.
Cuidado emocional	¿La evaluación preserva la confianza, la autonomía y la motivación?	Evitar vigilancia excesiva, lenguaje sancionador y retroalimentación despersonalizada.

Fuente: elaboración propia con base en UNESCO (2022), Miao y Holmes (2023), NIST (2023), UNICEF (2021), Floridi et al. (2018) y Hernández Armenta y Rosendo Olivares (2024).

A partir de los hallazgos anteriores, se propone un modelo de implementación en cinco fases. La primera fase consiste en el diagnóstico de pertinencia, donde la institución debe justificar el objetivo pedagógico del sistema. La segunda fase corresponde a la evaluación de riesgos, con énfasis en sesgos, privacidad, afectaciones emocionales y posibles impactos discriminatorios. La tercera fase es la implementación controlada, con capacitación docente, comunicación a estudiantes y mecanismos de consentimiento o información clara. La cuarta fase es la revisión humana y atención de inconformidades. La quinta fase es la auditoría periódica, orientada a corregir errores, actualizar criterios y evaluar consecuencias no previstas.

Tabla 3

Fases institucionales para implementar evaluación asistida por inteligencia artificial

Fase	Acción institucional	Resultado esperado
1. Diagnóstico de pertinencia	Definir finalidad educativa, población involucrada y necesidad real de la herramienta.	Evitar adopciones tecnológicas improvisadas o innecesarias.
2. Evaluación de impacto	Revisar sesgos, datos tratados, riesgos emocionales, accesibilidad y efectos en derechos.	Identificar daños previsibles antes del despliegue.
3. Implementación controlada	Capacitar al profesorado, informar al alumnado y establecer reglas de uso.	Asegurar comprensión, confianza y uso pedagógico responsable.
4. Revisión e impugnación	Habilitar procedimientos para solicitar explicación, corrección o revisión humana.	Proteger debido proceso académico y evitar decisiones automatizadas irrevisables.
5. Auditoría y mejora continua	Documentar errores, evaluar impactos diferenciados y actualizar protocolos.	Convertir la experiencia institucional en aprendizaje y prevención de daños.

Fuente: elaboración propia con base en OECD (2023), NIST (2023), European Parliament and Council of the European Union (2024) y UNESCO (2024a, 2024b).

Discusión

Los resultados permiten sostener que la pertinencia de la inteligencia artificial en la evaluación educativa no puede evaluarse solo por su capacidad técnica. Una herramienta puede corregir rápido, generar reportes y procesar grandes volúmenes de datos; sin embargo, la calidad educativa depende de que esos procesos sean válidos, justos, comprensibles y emocionalmente cuidadosos. Esta idea coincide con la tradición de la evaluación formativa, que entiende la retroalimentación como mediación para aprender y no como simple clasificación (Black & Wiliam, 1998; Nicol & Macfarlane-Dick, 2006).

La noción de vínculo simbiótico permite explicar por qué la inteligencia artificial no es un accesorio externo de la evaluación. Al integrarse a plataformas y decisiones institucionales, modifica qué datos se recogen, qué indicadores se consideran relevantes y qué actuaciones docentes se priorizan. De esta manera, la tecnología no solo ayuda a evaluar; también redefine la cultura evaluativa. Esta conclusión coincide con los planteamientos críticos que advierten que la inteligencia artificial en educación debe analizarse como fenómeno sociotécnico, atravesado por intereses, diseños, políticas, datos y relaciones de poder (Selwyn, 2019; Zawacki-Richter et al., 2019).

La discusión jurídica refuerza la necesidad de garantías. Cuando una recomendación automatizada afecta una calificación, una alerta de riesgo, la detección de plagio o la continuidad de una trayectoria escolar, se activa una exigencia mínima de explicabilidad, revisión e impugnación. La dignidad humana impide reducir al estudiante a un perfil estadístico. Asimismo, el derecho a la educación exige igualdad sustantiva y no discriminación, lo cual obliga a revisar impactos diferenciados en estudiantes con discapacidad, hablantes de lenguas distintas, personas neurodivergentes o estudiantes con brechas de conectividad (UNESCO, 1960; United Nations, 1948; United Nations, 2006). Cuando la herramienta se utiliza con niñas, niños o adolescentes, también debe considerarse el interés superior de la niñez y la protección reforzada frente a decisiones automatizadas que afecten su desarrollo educativo (United Nations, 1989).

Rosendo Olivares (2025) sostiene que los sesgos algorítmicos en sistemas de inteligencia artificial no deben entenderse como fallas menores o meramente técnicas, sino como riesgos capaces de reproducir desigualdades estructurales y afectar derechos humanos como la igualdad, la privacidad y el debido proceso. Esta lectura resulta especialmente pertinente para la evaluación educativa asistida por IA, ya que las decisiones automatizadas pueden incidir en calificaciones, trayectorias escolares, permanencia, becas o reputación académica del estudiantado.

El Reglamento de Inteligencia Artificial de la Unión Europea ofrece un referente comparado relevante porque clasifica ciertas aplicaciones educativas como sistemas de alto riesgo cuando inciden en acceso, admisión, evaluación o asignación de personas en instituciones educativas (European Parliament and Council of the European Union, 2024). Aunque su aplicación directa

corresponde al espacio europeo, su lógica de gestión basada en riesgo resulta útil para contextos latinoamericanos: cuanto mayor sea el impacto de una herramienta sobre la trayectoria de una persona estudiante, mayores deben ser las obligaciones de transparencia, documentación, supervisión y revisión.

La dimensión psicoemocional exige ampliar el debate. La evaluación automatizada puede producir ansiedad cuando el estudiante no comprende el criterio aplicado, cuando siente vigilancia permanente o cuando no puede dialogar con una persona responsable. Desde la teoría de la autodeterminación, el aprendizaje requiere experiencias de autonomía, competencia y vínculo; por ello, una evaluación tecnológica que se perciba como control absoluto puede debilitar la motivación y la confianza (Deci & Ryan, 2000). En este sentido, el bienestar emocional no es un elemento accesorio, sino una condición de legitimidad educativa.

El papel docente se redefine, pero no desaparece. La inteligencia artificial puede liberar tiempo, ofrecer diagnósticos y apoyar retroalimentación; sin embargo, el profesorado conserva la responsabilidad de interpretar, contextualizar y decidir pedagógicamente. La supervisión humana significativa no debe entenderse como una revisión formal o automática, sino como la posibilidad real de comprender el resultado, contrastarlo con evidencia diversa y corregirlo. De lo contrario, el docente corre el riesgo de convertirse en simple ejecutor de recomendaciones producidas por sistemas opacos.

La aportación de Hernández Armenta y Rosendo Olivares (2024) resulta pertinente porque permite trasladar principios de ética jurídica de la inteligencia artificial al campo educativo. Transparencia, justicia, equidad, responsabilidad, privacidad y confidencialidad no son principios exclusivos de la investigación jurídica; también son condiciones necesarias para una evaluación educativa legítima. La evaluación con IA requiere, por tanto, un puente entre ética tecnológica, derechos humanos y pedagogía.

Una limitación del presente estudio consiste en que sus resultados derivan de análisis documental y no de trabajo empírico con docentes o estudiantes. No obstante, esta limitación abre una línea futura de investigación: aplicar la matriz propuesta en instituciones concretas para evaluar percepciones, prácticas, riesgos y condiciones de gobernanza. También sería pertinente desarrollar instrumentos de medición sobre bienestar emocional frente a evaluación

automatizada, así como estudios comparados entre universidades con distintos niveles de regulación tecnológica.

Conclusiones

El análisis permite concluir que la inteligencia artificial puede ser una herramienta valiosa para la evaluación educativa, siempre que opere como apoyo y no como sustituto del juicio docente. Su uso responsable exige reconocer que evaluar implica interpretar aprendizajes, contextos y necesidades, no solo procesar datos o emitir calificaciones. La tecnología puede fortalecer la retroalimentación, personalizar apoyos y detectar rezagos, pero también puede reproducir sesgos, afectar derechos y deteriorar el bienestar emocional si se implementa sin criterios éticos y jurídicos.

El vínculo entre inteligencia artificial y evaluación educativa es simbiótico porque ambas dimensiones se transforman mutuamente. La evaluación alimenta a la inteligencia artificial con datos y criterios; la inteligencia artificial reconfigura la forma en que se observa, clasifica y decide sobre el aprendizaje. Esta relación no es neutral. Puede fortalecer la justicia educativa cuando se gobierna con transparencia, equidad y revisión humana; pero puede profundizar desigualdades cuando opera como caja negra, con criterios opacos y sin posibilidad de impugnación.

Los derechos humanos ofrecen un límite indispensable. La dignidad, la igualdad, la no discriminación, la privacidad, la protección de datos, el derecho a la educación y el debido proceso académico deben orientar cualquier sistema de evaluación asistido por inteligencia artificial. En consecuencia, las instituciones educativas deben justificar la finalidad de uso, evaluar riesgos, informar al alumnado, capacitar al profesorado, documentar decisiones, habilitar revisión humana y realizar auditorías periódicas.

El bienestar emocional del estudiante debe incorporarse como criterio de calidad educativa. Una evaluación técnicamente eficiente pero emocionalmente dañina no puede considerarse plenamente legítima. La ansiedad, la sensación de vigilancia, la desconfianza o la pérdida de autonomía son efectos que deben prevenirse mediante comunicación clara, acompañamiento docente, límites de monitoreo y retroalimentación comprensible.

Como aportación principal, el artículo propone una matriz de principios para la evaluación asistida por inteligencia artificial basada en finalidad pedagógica legítima, proporcionalidad, transparencia, explicabilidad, trazabilidad, equidad, revisión humana significativa, protección de datos y cuidado emocional. Estos criterios pueden orientar políticas institucionales, protocolos docentes, comités de ética y futuras investigaciones empíricas. En síntesis, la pregunta educativa no debe ser únicamente qué puede hacer la inteligencia artificial, sino bajo qué condiciones debe utilizarse para respetar la dignidad humana, fortalecer el aprendizaje y evitar nuevas formas de exclusión.

Referencias bibliográficas

- Baker, R. S., & Hawn, A. (2022). Algorithmic bias in education. *International Journal of Artificial Intelligence in Education*, 32(4), 1052-1092. <https://doi.org/10.1007/s40593-021-00285-9>
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, 5(1), 7-74. <https://doi.org/10.1080/0969595980050102>
- Deci, E. L., & Ryan, R. M. (2000). The what and why of goal pursuits: Human needs and the self-determination of behavior. *Psychological Inquiry*, 11(4), 227-268. https://doi.org/10.1207/S15327965PLI1104_01
- European Parliament and Council of the European Union. (2024). Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). *Official Journal of the European Union*. <https://eur-lex.europa.eu/eli/reg/2024/1689/oj>
- Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., & Vayena, E. (2018). AI4People-An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. *Minds and Machines*, 28, 689-707. <https://doi.org/10.1007/s11023-018-9482-5>

-
- High-Level Expert Group on Artificial Intelligence. (2019). Ethics guidelines for trustworthy AI. European Commission. <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>
- Hernández Armenta, C., & Rosendo Olivares, F. (2024). Aspectos éticos en la aplicación de inteligencia artificial para la investigación jurídica. *Pensamiento Crítico. Revista de Investigación Multidisciplinaria*, 10(19), 1-14. <https://doi.org/10.64040/nyyryn02>
- Holmes, W., Bialik, M., & Fadel, C. (2019). Artificial intelligence in education: Promises and implications for teaching and learning. Center for Curriculum Redesign.
- Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, 50(9), 741-749. <https://doi.org/10.1037/0003-066X.50.9.741>
- Miao, F., & Holmes, W. (2023). Guidance for generative AI in education and research. UNESCO. <https://unesdoc.unesco.org/ark:/48223/pf0000386693>
- National Institute of Standards and Technology. (2023). Artificial Intelligence Risk Management Framework (AI RMF 1.0). U.S. Department of Commerce. <https://doi.org/10.6028/NIST.AI.100-1>
- Nicol, D. J., & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in Higher Education*, 31(2), 199-218. <https://doi.org/10.1080/03075070600572090>
- Organisation for Economic Co-operation and Development. (2023). OECD Digital Education Outlook 2023: Towards an effective digital education ecosystem. OECD Publishing. <https://doi.org/10.1787/c74f03de-en>
- Popenici, S. A. D., & Kerr, S. (2017). Exploring the impact of artificial intelligence on teaching and learning in higher education. *Research and Practice in Technology Enhanced Learning*, 12, Article 22. <https://doi.org/10.1186/s41039-017-0062-8>

-
- Rosendo Olivares, F. (2025). Sesgos algorítmicos en sistemas de inteligencias artificiales: Riesgo para los derechos humanos y desafíos jurídicos. *Contacto IUS*, 13, 28-39. https://www.researchgate.net/publication/396437184_Sesgos_algoritmicos_en_sistemas_de_inteligencia_artificiales_para_los_derechos_humanos_y_desafios_juridicos
- Selwyn, N. (2019). *Should robots replace teachers? AI and the future of education*. Polity Press.
- Slade, S., & Prinsloo, P. (2013). Learning analytics: Ethical issues and dilemmas. *American Behavioral Scientist*, 57(10), 1510-1529. <https://doi.org/10.1177/0002764213479366>
- Tang, X., Upadyaya, K., Toyama, H., Kasanen, M., & Salmela-Aro, K. (2023). Assessing and tracking students' wellbeing through an automated scoring system: School Day Wellbeing Model. In H. Niemi, R. D. Pea, & Y. Lu (Eds.), *AI in learning: Designing the future* (pp. 55-71). Springer. https://doi.org/10.1007/978-3-031-09687-7_4
- UNESCO. (1960). *Convention against discrimination in education*. <https://unesdoc.unesco.org/ark:/48223/pf0000183342>
- UNESCO. (2022). *Recommendation on the ethics of artificial intelligence*. <https://unesdoc.unesco.org/ark:/48223/pf0000380455>
- UNESCO. (2024a). *AI competency framework for teachers*. <https://unesdoc.unesco.org/ark:/48223/pf0000391104>
- UNESCO. (2024b). *AI competency framework for students*. <https://unesdoc.unesco.org/ark:/48223/pf0000391105>
- UNICEF. (2021). *Policy guidance on AI for children*. UNICEF Office of Global Insight and Policy. <https://www.unicef.org/innocenti/reports/policy-guidance-ai-children>
- United Nations. (1948). *Universal Declaration of Human Rights*. <https://www.un.org/en/about-us/universal-declaration-of-human-rights>

United Nations. (1989). Convention on the Rights of the Child.
<https://www.ohchr.org/en/instruments-mechanisms/instruments/convention-rights-child>

United Nations. (2006). Convention on the Rights of Persons with Disabilities.
<https://www.ohchr.org/en/instruments-mechanisms/instruments/convention-rights-persons-disabilities>

Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education: Where are the educators? *International Journal of Educational Technology in Higher Education*, 16, Article 39. <https://doi.org/10.1186/s41239-019-0171-0>

Declaraciones finales

Conflicto de intereses: Los autores declaran que no existe conflicto de interés posible.

Financiamiento: No existió asistencia financiera de partes externas al presente artículo.

Agradecimiento: N/A

Nota editorial: El artículo no es producto de una publicación anterior ni se encuentra sometido simultáneamente a otra revista.